University of Bedfordshire

Cognitive validity

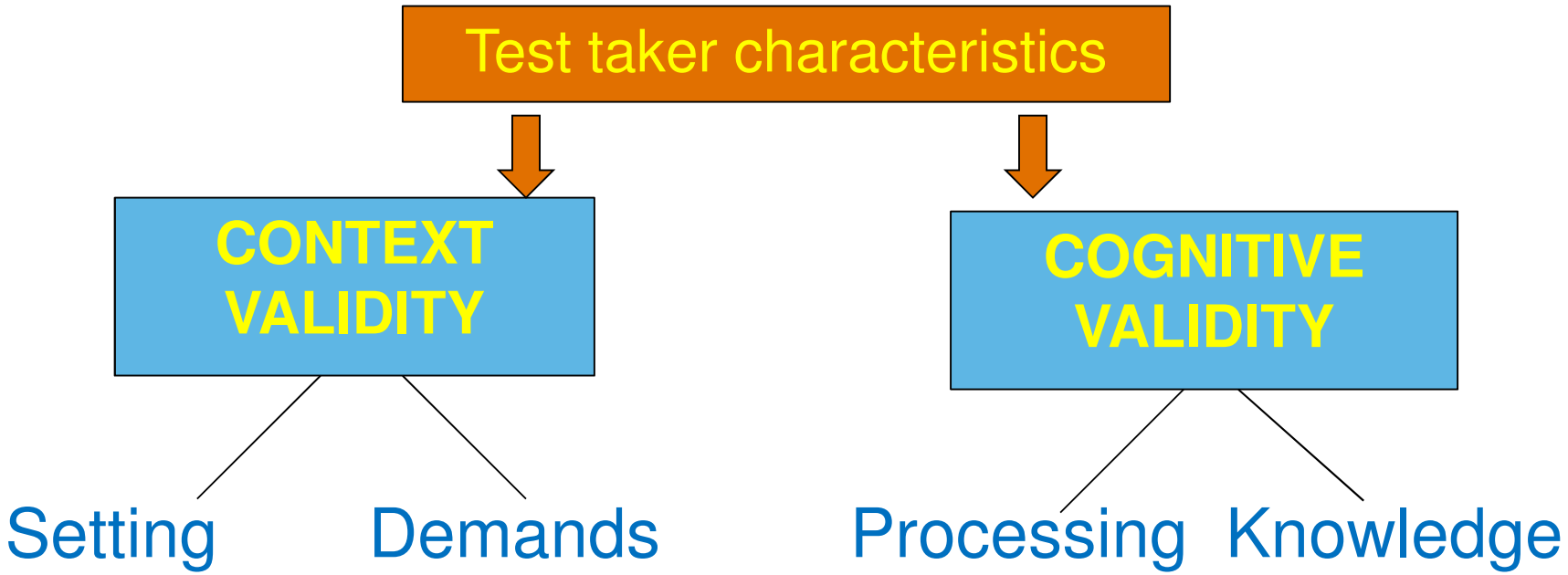
Dr John Field, CRELLA

University of Bedfordshire

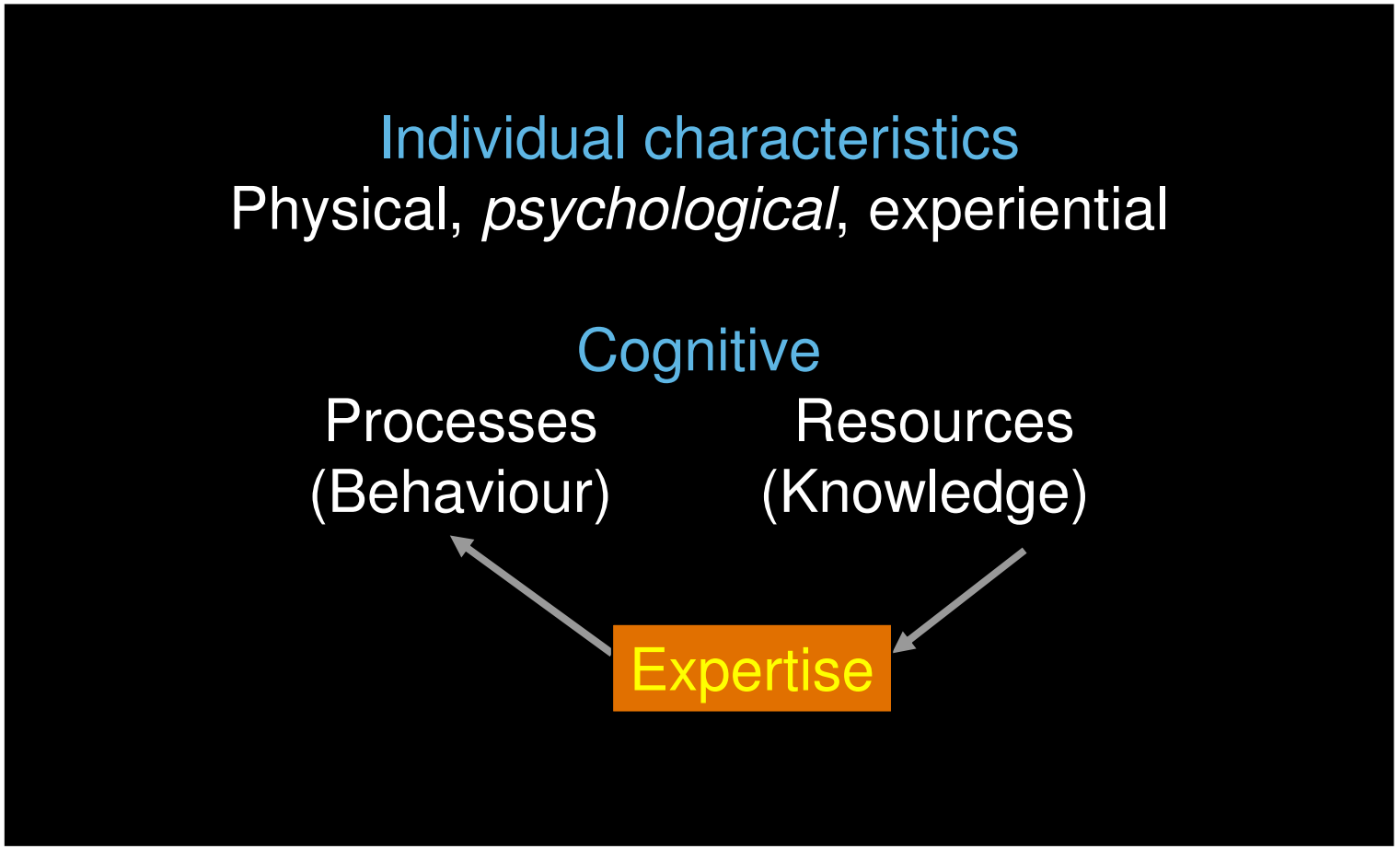
Background

The socio-cognitive framework (Weir, 2005)

University of Bedfordshire



The test taker (O'Sullivan, 2011)



An expert

University of Bedfordshire

- is somebody who can use a skill in way that is rapid and that does not demand a great deal of forethought.
- A good driver does not have to think about the process of changing gears.
- A good speaker constructs and produces a sentence without having to pause to think about the words or grammar being used.
- A test of a language skill should not assess knowledge of language, but expertise in using that language.

Predictive testing

University of Bedfordshire

- Many high stakes language test scores are employed predictively: e.g. to show that an individual is capable of performing in a particular job, class or academic setting.
- This places a responsibility on the test designer to ensure that the test elicits behaviour similar to the behaviour that happens in a real-world context.

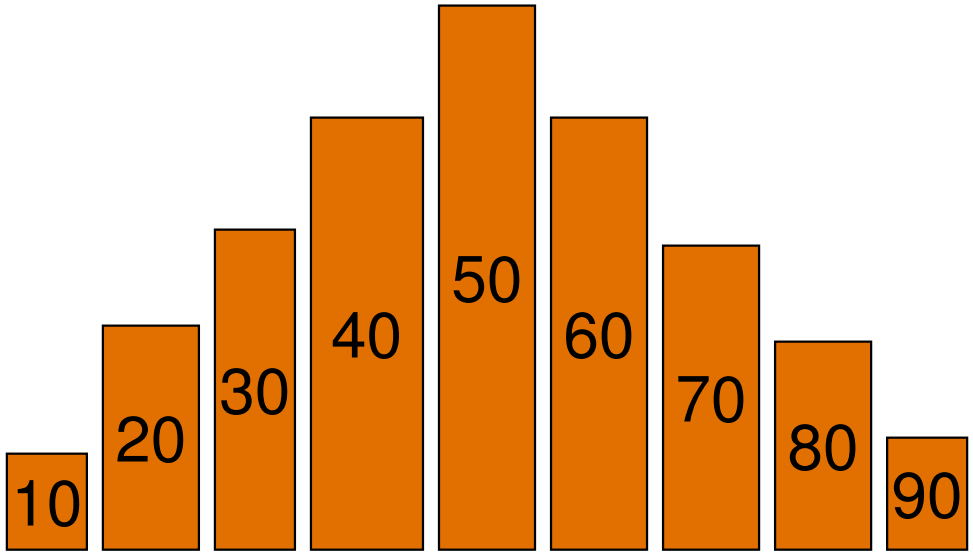
Cognitive validity (Glaser, 1991)

University of Bedfordshire

- Clearly we cannot reproduce the circumstances of a real listening event in the artificial environment of a test.
- But cognitive validity requires us to find out if the mental processes that a test elicits from a candidate resemble the processes that he/she would employ in non-test conditions.
- At issue: *How valid is the test as a predictor of real-life performance?*
- The notion of cognitive validity has been used to investigate whether tests of scientific thinking or logical reasoning actually tap in to the processes they are supposed to measure (rather than, e.g. relying on rote learned facts). Baxter & Glaser, 1998, Thelk & Hoole, 2006

Distribution

University of Bedfordshire



Construct validation

University of Bedfordshire

- These hypothetical results from piloting a test show that it discriminates well between learners. It might seem to test the construct (say listening) well.
- *But supposing it is a test of listening with an easy recording but very complex comprehension items?*
- It is possible that the result mainly represents learners' reading skills and not their listening skills

Establishing cognitive validity

University of Bedfordshire

- Weir (2005) argues that we need to have a clearer idea of the construct we are testing before designing a test.
- Cognitive validity can be investigated in two ways:

1. How does an expert listener behave (what is the target behaviour learners are working towards?)

Modelling the skill

2. What do test takers actually do in a listening test? How closely does it resemble natural listening?

Studying candidate behaviour (verbal report)

University of Bedfordshire

Modelling the skill

Evidence for listening

University of Bedfordshire

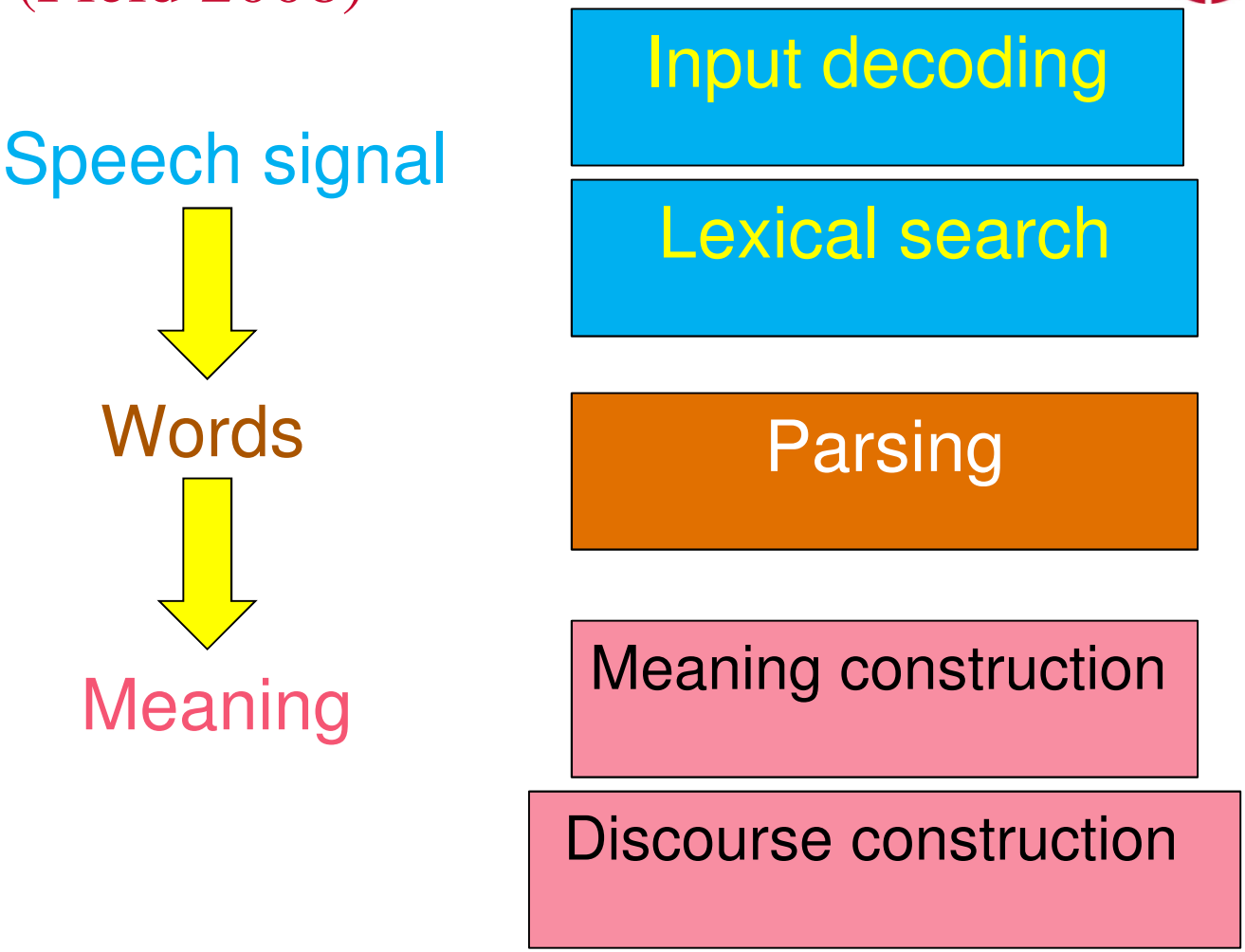
- Testers and researchers need information about
  - the speech signal that reaches a listener's ear, and the problems it might cause to an L2 listener. [Source: phonetics]
  - the processes that an expert listener uses in normal circumstances and the way they might vary in the case of an L2 listener. [Source: psycholinguistics]

The speech signal

University of Bedfordshire

- Speech is highly variable at phoneme, word and phrase level (cf spelling)
- There are no consistent gaps between words
- Speech is transitory: a listener cannot go back to check
- Listening is time-constrained. A listener cannot speed up or slow down ,
- Difficulty in listening is partly determined by speaker factors: voice - speech rate - precision of articulation - hesitation conversational features - accent - rhythmicity

Five phases of listening
(Field 2008)



Three cognitive validity questions

University of Bedfordshire

- 1. To what extent are the cognitive processes elicited by a test comparable to those that would be employed in a real-world setting?
- 2. Is the range of processes elicited by a test comprehensive enough to be representative of behaviour in a real-world setting?
- 3. Are the cognitive demands imposed by a test sufficiently finely calibrated to reflect the level of the test?

Listening test components that may compromise cognitive validity

University of Bedfordshire

- Recording
- Recording-as-text
- Test format
- Items

Test format conventions

University of Bedfordshire

- Items are presented before listening. They
  - provide more information than would normally be available ahead of listening (and provide it in a different modality)
  - encourage the candidate to anticipate what will be heard (sometimes incorrectly)
- The need to read and internalise the items dictates that
  - items have to be presented in the same order as the passage.
  - Items have to be spaced out

Sample format: MCQ

University of Bedfordshire

You hear a man and a woman talking about going to the gym. What does the man say about going to the gym?

A. It is too expensive for him

B. It takes too much of his time.

C. It is too physically demanding

(*FCE Handbook*, 2008: 68)

Recording at text 2

University of Bedfordshire

Woman: So that didn't last long, did it? Two weeks going to the gym and you're already talking about giving it up…

Man: Look, if you're saying I'm not up to it, you're wrong. I realise it's very effective in working every muscle, and when I get started, it's just like other sports. I don't even mind feeling exhausted at the end. But, listen, you sort out your kit at home, lug it to the gym, queue to pay your entrance fee, then change and queue for the machines … when you could have been for a run straight from your home and then been free to get on with your life.

Woman: Well, I think you're wrong and you should make the effort to carry on.

Recording as text 2

University of Bedfordshire

Woman: So that didn't last long, did it? **Two weeks going to the gym** and you're already **talking about giving it up…**

Man: Look**, if you're saying I'm not up to it, you're wrong**. I realise **it's very effective in working every muscle**, and when I get started, **it's just like other sports**. **I don't even mind feeling exhausted** at the end. But, listen, **you sort out your kit at home, lug it to the gym**, **queue to pay your entrance fee**, then **change and queue for the machines** … when **you could have been for a run straight from your home** and **then been free to get on with your life.**

Woman: Well, **I think you're wrong** and **you should make the effort to carry on.**

Recording as text 2

University of Bedfordshire

Woman: So that didn't last long, did it? Two weeks going to the gym and you're already talking about giving it up…

Man: Look, if you're saying I'm not up to it, you're wrong. I realise it's very effective in working every muscle, and when I get started, it's just like other sports. I don't even mind feeling exhausted at the end. But, listen, *you sort out your kit at home, lug it to the gym, queue to pay your entrance fee, then change and queue for the machines …* when *you could have been for a run straight from your home* and then been free to get on with your life.

Woman: Well, *I think you're wrong* and you should make the effort to carry on.

Recording as text 2

University of Bedfordshire

- Test setters tend to base their tests on a written script which has not yet been recorded.
- The linguistic criteria they employ rely heavily on lexical frequency and syntactic simplicity.
- BUT in processing terms difficulty is often caused by:
- a. the density of ideas and the complexity of the links between them
- b. perceptual saliency of phrases and clauses

Processing of test formats

University of Bedfordshire

- The test formats used in listening tests are chosen because of their reliability and ease of marking. BUT
- They impose quite heavy cognitive demands upon the candidate who has to:
  - Internalise information from the items
  - Map from the items to the listening passage (which they often paraphrase)
  - Decide how closely each new idea in the listening passage fits the current item
  - (MCQ) eliminate options that are negated in the recording.
- These operations are more demanding than normal listening

Sample phase:
Discourse construction

University of Bedfordshire

Choose	**Is it important? Is it relevant?**
Connect	**How is it linked to the last utterance?**
Compare	**Is it consistent with what was said so far?**
Construct	**What is the overall line of argument?**

Discourse building overlooked

University of Bedfordshire

- Choose: the tester chooses which information points to focus on – sometimes choosing points that are not central to the recording
- Connect: Much testing focuses on single points, with no connection to those before and after
- Compare: Tests rarely ask learners to check information (for example, comparing two accounts of an accident)
- Construct. Tests rarely seek for evidence that learners can construct an outline based upon macro-and micro points / headings and subheadings

University of Bedfordshire

Examining candidate behaviour

Three types of behaviour (Field, 2009)

University of Bedfordshire

- 1 Part of the normal process: behaviour which might be adopted by an L1 academic listener.
- 2. Strategic behaviour to prepare for a task, to maximise the amount retained or to compensate for problems of understanding.
- 3 Task-specific behaviour representing the user's response to features of the task.
    - a. processes related to the task but not part of the corresponding real-life activity
    - b. strategies where the learner attempts to exploit loopholes in the format of the task

Academic cognitive validity study (Field, 2009)

University of Bedfordshire

- Academic listening test played to participants under test conditions
- Recording paused three times to ask for answers and elicit reasons for choosing the answers.
- Second academic listening test played to participants under lecture conditions (note-taking and reporting back what they had understood).

Task-specific behaviour

The items in a gap-filling task potentially provide a candidate with:

- An outline of what the recording covers
- A set of gaps to be filled that follow the sequence of the recording
- Key words with which to locate information
- One constituent of a collocation
- Sequences which echo the recording word for word

Attentional focus

- (1) [The main point was] preserve tree but I'm not quite sure because + every every time I use + I mean my my method to + listen to to do the […] listening + yeah I just look at the words not focus what it is about (S1: 145)
- (2) So when I was reading the answering the first one she was maybe she had already finished the list no? + the other case is even if the words maybe were I made some mistakes in other parts I mean + but you have time to write to listen because when you were when I was writing er she was speaking about something else not important for the test. (AA1: 148)

Unfortunate consequences

- (3) I missed it because I didn't I didn't I didn't realise the 'frequency' has came so quickly (P1: 99)
- (4) Er when I try to get this answer um he he is already talking about the make cities cooler yes so I missed the answer (T1: 19)

Task-wise strategies

University of Bedfordshire

- Use words from written text to locate information
- Listen for one-to-one match
- Seek paraphrase for a written proposition
- Choose answer according to position in a list or sequence in the *written* text

The ‘key word’ strategy

University of Bedfordshire

- Instructors encourage learners to exploit the written material in a listening task by listening out for key words identified from the questions.
- The evidence suggests that this can be counter-productive. If a learner misses a given key word match, there can be a knock-on effect. The learner’s attention continues to focus on the target and misses subsequent ones (Field, 2009).

Perceived difficulty

University of Bedfordshire

- Hypothesis: Task condition will be found to be easier than Lecture condition.
- 30% of Participants reported that they found note-taking in the lecture easier than doing the task

Typical comment favouring Lecture (Field, 2009)

Which did you find more difficult?

The first one + definitely. […] Because um with this this test it it is um + how can I? It's + I have I have a lot of more stress with this sort of test because you um you don't want to miss any answer + but with this technique it's it is different + um even if you miss something er you you you will understand the the general idea what is talking about + but in this test sort of test um you when you miss miss the point then you get you get stress and then for er for the following questions it's harder and so it's quite difficult.

… were you listening in a different way?

Um yes definitely +there's not the same way + yeah +because in the first part I'm just focusing on words not the general meaning + but in the second I focus on the the the general ideas + the most interesting points. (V2: 26)

Conclusion

Cognitive Validity research:

- 1. compares what we know from empirical findings about the processes that contribute to a target construct

  against the input to the test taker

- against the formats used

  against item content

- 2. compares the behaviour of the test taker under test conditions and under those that more closely replicate real world language use.

References

- Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press
- Field, J. (2009). The cognitive validity of the lecture listening section of the IELTS listening paper. *IELTS Research Reports 9*, Cambridge ESOL
- Field, J. (2011). Cognitive validity. In Taylor, L. (ed.) *Examining Speaking*. Cambridge, Cambridge University Press
- Field, J. (forthcoming). Cognitive validity. In Geranpayeh. A. & Taylor, L. (eds.) *Examining Listenng*. Cambridge, Cambridge University Press
- Glaser, R., 1991. 'Expertise and assessment'. In M.C. Wittrock and E.L. Baker (eds.) *Testing and Cognition*. Englewood Cliffs, Prentice Hall, pp 17-30
- Weir, C., 2005. *Language testing and validation: an evidence-based approach.* Basingstoke, Palgrave Macmillan

John.Field@beds.ac.uk